Boundary Ownership by Lifting to 2.1D

Ido Leichter and Michael Lindenbaum
Technion – Israel Institute of Technology
Computer Science Department, Technion, Haifa 32000, Israel
{idol,mic}@cs.technion.ac.il

Abstract

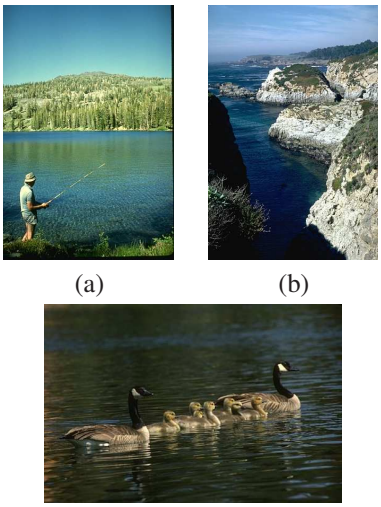
This paper addresses the "boundary ownership" problem, also known as the figure/ground assignment problem. Estimating boundary ownerships is a key step in perceptual organization: it allows higher-level processing to be applied on non-accidental shapes corresponding to figural regions. Existing methods for estimating the boundary ownerships for a given set of boundary curves model the probability distribution function (PDF) of the binary figure/ground random variables associated with the curves. Instead of modeling this PDF directly, the proposed method uses the 2.1D model: it models the PDF of the ordinal depths of the image segments enclosed by the curves. After this PDF is maximized, the boundary ownership of a curve is determined according to the ordinal depths of the two image segments it abuts. This method has two advantages: first, boundary ownership configurations inconsistent with every depth ordering (and thus very likely to be incorrect) are eliminated from consideration; second, it allows for the integration of cues related to image segments (not necessarily adjacent) in addition to those related to the curves. The proposed method models the PDF as a conditional random field (CRF) conditioned on cues related to the curves, T-junctions, and image segments. The CRF is formulated using learnt non-parametric distributions of the cues. The method significantly improves the currently achieved figure/ground assignment accuracy, with 20.7% fewer errors in the Berkeley Segmentation Dataset.

1. Introduction

The "boundary ownership" problem is the problem of automatically deciding which of the two image segments abutting a provided curve along an object's boundary in an image is the 'figure' (i.e., "owns the boundary") and which is the 'ground'. Although the image segment corresponding to the occluding object is usually the figure, boundary ownership is generally a perceptual characteristic. That is, the image region that is perceived as lying in front of the ground regions is considered as 'figure' [11], even if this perception contradicts the 3D layout of the scene. Examples of this may be seen in the images in Fig. 1, taken from the Berkeley Segmentation Dataset (BSDS). They were first segmented by a human observer, after which two additional human observers attributed each of the boundary curves between the segments to one of its two abutting segments. Although the water in images (a) and (b) occludes the land and the rocks beneath it, both human observers marked the land and the rocks as the "owners" of the shorelines. (This may explain the name 'shoreline' rather than, say, 'sealine'.) Another example is provided in image (c), where both human observers marked the ducks as the owners of the contact line between the ducks and the water, although the water occludes the duck parts below this contact line. A different type of example is the occasional perception of a hole as the figure in cases where the background seen through the hole does not appear in the image around the object containing the hole [11].

It has been shown that boundary ownership perception is influenced by many factors, including local characteristics of the boundary shape (e.g., convexity [6] and orientation, termed "lower region" [16]), its global characteristics (e.g., symmetry, size and orientation [6, 11]), the appearance of the region itself (e.g., contrast [11]) and also its surroundings (e.g., parallelism and surroundedness [11]). The orientations of the intensity level sets in the close neighborhood of the boundary have been shown to bear information about which of the two boundary sides occludes the continuation of the other. (The orientations of the intensity level sets are measured relative to that of the boundary itself.) This information may thus be used for estimating the boundary ownership as well [5].

All methods for automatic figure/ground assignment in images measure a subset of the above factors. These measurements, which are usually ad hoc, are used to make the boundary ownership decisions. In some works, these measurements are performed for each boundary pixel by analyzing the boundary curve segment and the image region in its neighborhood. The decision for each boundary pixel



(a) (b) (c)

Figure 1. Cases where the boundary side corresponding to the occluding object was not classified as 'figure' by the human observers. See text for details.

is then made independently. For example, such an approach was taken in [5], where the orientations of the intensity level sets near the boundary were used for deciding which of the boundary sides occludes the continuation of the other. Another example is [1], where the convexity, lower region and size cues were locally measured for boundary points and then used to independently estimate the figure/ground label at each point. In one of the methods in [15], the assignment for each boundary pixel was also made independently, according to the likelihood of a set of local shape descriptors that was based on a reference set of learnt *shapemes* [9]. These shapemes are prototypical shapes that implicitly encode cues such as convexity and parallelism.

The figure/ground assignment along a boundary curve between two junctions is constant. This suggests that better precision can be obtained by using cues related to the whole curve segment and making a curve-based, rather than a pixel-based, figure/ground assignment. Such an approach was taken, for example, in [13] and in [15]. In fact, in [15] the figure/ground assignment to the curves was made by averaging the local shapeme on the curve. The latter method provided significantly more accurate results than the local model and reinforced the advantage of making curve-based assignments.

Curve junctions and the angles between the curves at the junctions provide additional cues for the figure/ground labels of the junction curves. Thus, incorporating this additional information in the boundary ownership estimates might increase the precision further. Since each junction provides a cue for the mutual figure/ground labels of all its curves (rather than on each curve independently), the integration of the junction cues into the boundary ownership model causes the curve assignments to be coupled. Curve junctions in figure/ground labeling were used, for example, in [13] and [15]. Both these works used a conditional random field (CRF) [8] to model the probability distribution function (PDF) of the random variables associated with the curve labels. In fact, in [15], the CRF model yielded significantly more accurate results than the model that did not use the junction cues, reinforcing their importance. Modeling the distribution of the sought labels as a CRF was also done in related problems such as contour completion [14] and the detection of occlusion boundaries [4].

Instead of estimating the figure/ground labels of the boundary curves directly, the proposed method uses the 2.1D model [10] to model the PDF of the ordinal depths of the image segments enclosed by the curves. The image is partitioned into the segments enclosed by the provided curves, the depicted scene is approximated as 2.1D, and the ordinal depths of these segments are estimated by maximizing a PDF of these depths. Then, the boundary ownership of each curve is decided from the ordinal depths of the two image segments abutting it: the curve side consisting of the image segment that is "in front" of the image segment on the other curve side is estimated as the owner of the curve, i.e., the figure. Although the 2.1D model is not applicable for all images in their entirety, it is applicable for most natural images and is typically a reasonable approximation for the rest. Note that any inconsistency between the 2.1D model and the actual 3D scene depicted in the image is not important: as demonstrated in Fig. 1, the perceived layer ordering is what counts.

Estimating the ordinal depths of the image segments has two advantages over estimating the figure/ground labels directly. The first advantage is that boundary ownership configurations inconsistent with every depth ordering (and thus very likely to be incorrect) are eliminated. An example where the 2.1D model is essential is provided in Fig. 2. In this figure, it is impossible to deduce the ownership of the marked curve by considering the curves and their junctions alone. Using the 2.1D model will succeed simply because the wrong (impossible) choice will not be considered. The second advantage is the ability to integrate cues related to image segments in addition to those related to the curves. Moreover, the segment cues may even relate non-adjacent segments. These two advantages could increase the accuracy of the figure/ground assignment. The proposed method models the PDF of the segments' ordinal depths as a conditional random field (CRF) conditioned on cues related to the curves, T-junctions, and image segments. The cues are integrated in the CRF by using their conditional likelihoods (conditional on the curve labels), where these likelihoods are determined from learnt non-parametric PDFs of the cues. Experiments using the BSDS show that the

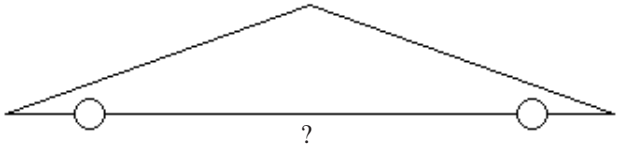


Figure 2. It is impossible to deduce the ownership of the marked curve by considering the curves and their junctions alone.

method significantly improves on the current figure/ground assignment accuracy.

A related work is [13], where a CRF was used to model the binary figure/ground labels of image regions in images consisting of one figure of a specific class (horses). Another related work is [4], where the occlusion boundaries are recovered as well (in addition to their figure/ground assignments). Note that the depth ranges of the image segments in the latter work are estimated after the boundary ownerships are determined, whereas here it is the other way around.

The rest of the paper proceeds as follows: Sec. 2 provides an outline of the proposed method. In Sec. 3 the probabilistic model is described in detail. Experiments and their results are provided in Sec. 4, and Sec. 5 concludes the paper with a discussion.

2. Algorithm outline

An image and a set of curves marked along object boundaries are provided. The goal is to perform, for each curve C_i, $i = 1, \ldots, N$, one of the following two label assignments: $L_i = \text{FG}$ (foreground) – the right hand-side of the curve is the 'figure', or $L_i = \text{BG}$ (background) – the right hand-side of the curve is the 'ground'. Note that the assignment of the curve sides to 'right' and 'left' is determined by the direction of the curve, which is arbitrary.

We estimate the labels by the following steps:

1. The image is partitioned into the segments S_i, $i = 1, \ldots, M$, enclosed by the curves.

2. Using the 2.1D model, we construct a probability distribution function of the ordinal depths of these image segments.

3. The maximum a posteriori (MAP) ordinal depth configuration $d_i \in \{1, \ldots, M\}$, $i = 1, \ldots, M$, for the segments S_i is estimated ($d_i < d_j$ means S_i is in front of S_j).

4. Each curve is labeled according to which of its two abutting image segments is in front.

3. The probabilistic model

The PDF of the segments' ordinal depths is modeled as the Gibbs distribution

$$\Pr\left(\{d_i\}_{i=1}^{M}\right) \propto \exp\left\{-\left(\sum_{i=1}^{N} V_i\left(d_{i_{\text{left}}}, d_{i_{\text{right}}}\right) + \sum_{\text{T-junctions } y} V^y\left(d_{y_1}, d_{y_2}, d_{y_3}\right) + \sum_{\{j,k\}\in\mathcal{P}} V_{j,k}\left(d_j, d_k\right)\right)\right\}, \quad (1)$$

where i_{left} and i_{right} are the indices of the segments to the left and to the right of C_i, respectively; y_1, y_2 and y_3 denote the indices of the three segments sharing T-junction y; and $\mathcal{P}$ is a set of unordered non-adjacent segment pairs (to be defined in Sec. 3.3). This distribution is a conditional random field with respect to the neighborhood system consisting of all segment pairs separated by a curve, segment triplets sharing a T-junction of curves, and a subset of the pairs of non-adjacent segments [2, 8]. Each potential function V_i in the first sum models the depth relation between the two image segments abutting curve C_i by using cues related to this curve. Each potential function V^y in the second sum models the depth relations between the three image segments sharing T-junction y by using cues related to this junction. Each potential function $V_{j,k}$ in the third sum models the depth relation between the two non-adjacent segments S_j and S_k by using cues related to this segment pair. In the following we discuss these potential functions in detail.

3.1. Curve-related potential functions

The potential functions V_i are defined as the negative log posterior of the two segments' ordinal depths conditioned on the curve cues, weighted by the curve length $|C_i|$,

$$V_i(d_{i_{\text{left}}}, d_{i_{\text{right}}}) = -|C_i| \log \Pr\left(d_{i_{\text{left}}}, d_{i_{\text{right}}} | \text{cues of } C_i\right). \quad (2)$$

Note that by the term 'curve cues' we refer to all the geometric and photometric cues *associated* with the curve, and not only to those that depend merely on the curve itself. The curve cues only give an indication as to which of the two curve sides is in front of the other. Thus,

$$\Pr\left(d_{i_{\text{left}}}, d_{i_{\text{right}}} | \text{cues of } C_i\right) \propto \Pr\left(L_i | \text{cues of } C_i\right). \quad (3)$$

(The events $d_{i_{\text{left}}} > d_{i_{\text{right}}}$ and $d_{i_{\text{left}}} < d_{i_{\text{right}}}$ are associated with the events $L_i = \text{FG}$ and $L_i = \text{BG}$, respectively, and $\Pr\left(d_{i_{\text{left}}} = d_{i_{\text{right}}}\right) = 0$ by assumption.) Since the prior probabilities $\Pr\left(L_i = \text{FG}\right)$ and $\Pr\left(L_i = \text{BG}\right)$ are equal, the posterior is proportional to the cue likelihoods:

$$\Pr\left(L_i | \text{cues of } C_i\right) \propto p\left(\text{cues of } C_i | L_i\right). \quad (4)$$

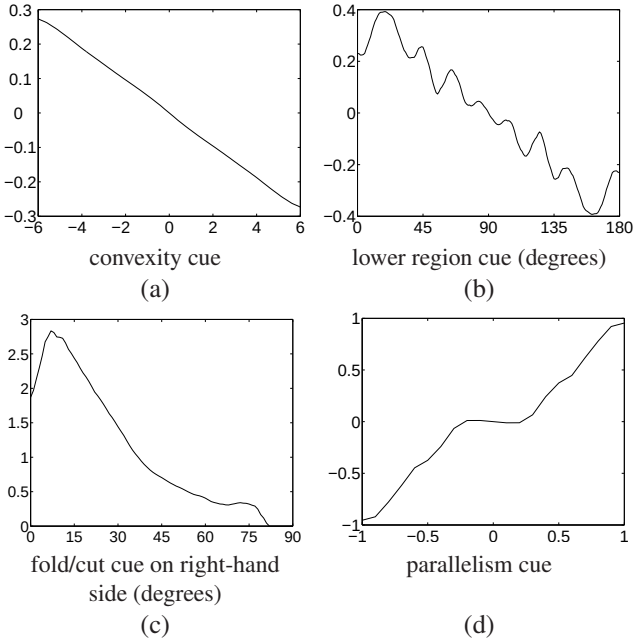


Figure 3. (a) $\log\left(p\left(\text{convexity cue}\middle|\text{FG}\right)/p\left(\text{convexity cue}\middle|\text{BG}\right)\right)$. (b) $\log\left(p\left(\text{lower region cue}\middle|\text{FG}\right)/p\left(\text{lower region cue}\middle|\text{BG}\right)\right)$. (c) $\dfrac{p\left(\text{fold/cut cue on right-hand side}\middle|\text{FG}\right)}{p\left(\text{fold/cut cue on right-hand side}\middle|\text{BG}\right)}$. (d) $\log\left(p\left(\text{parallelism cue}\middle|\text{FG}\right)/p\left(\text{parallelism cue}\middle|\text{BG}\right)\right)$. All the logarithms are to the base of 10.

Four different types of cues are measured for each curve, and these cues are approximated as conditionally independent (conditioned on the curve label):

$$\begin{aligned} p\left(\text{cues of } C_i\middle|L_i\right) &= p\left(\text{convexity cue of } C_i\middle|L_i\right) \\ &\cdot p\left(\text{lower region cue of } C_i\middle|L_i\right) \\ &\cdot p\left(\text{fold/cut cue of } C_i\middle|L_i\right) \\ &\cdot p\left(\text{parallelism cue of } C_i\middle|L_i\right). \end{aligned} \quad (5)$$

All these conditional distributions are non-parametric distributions learnt in a training phase from the BSDS by calculating the normalized (smoothed) histograms of the cues.

Convexity cue. The convexity cue is a geometric cue that is based on the tendency of the convex side of the curve to be perceived as the 'ground' side [11]. Therefore, we define the convexity cue to be the integral of the curvature along the curve. The log-ratio of the conditional likelihoods of the cue is plotted in Fig. 3(a). The expected tendency of the curve's convex side to be 'ground' is apparent.

Lower region cue. The lower region cue is a geometric cue that is based on the tendency of the lower side of the curve to be perceived as the figure side [1]. Therefore, we define it to be the mean of the absolute angle of the curve tangent ($-\pi \leq \text{angle} \leq \pi$). The log-ratio of the conditional likelihoods of the cue is plotted in Fig. 3(b). As expected, the curve's lower side tends to be perceived as figure.

Fold/cut cue. As explained in [5], the orientation of the intensity level sets along the close neighborhood of an occlusion boundary differs between the two boundary sides: The intensity level sets are tangent to the boundary at the boundary side that corresponds to a smooth occluding object ("fold"), and they are transverse to the boundary at the boundary side that corresponds to the obscured surface ("cut"). In accordance with this photometric phenomenon, we define the fold/cut cue as

$$\begin{aligned} &p\left(\text{fold/cut cue of } C_i\middle|L_i\right) \\ &\quad = p\left(\text{fold/cut cue on right-hand side of } C_i\middle|L_i\right) \\ &\qquad \cdot p\left(\text{fold/cut cue on left-hand side of } C_i\middle|L_i\right), \end{aligned} \quad (6)$$

where the fold/cut cue on a curve side is the weighted mean over all curve pixels p of the absolute difference between the angle of the curve normal at p and the angle of the intensity gradient near p. The mean is weighted by the intensity gradient size. The mean over the three color bands (RGB) is calculated. Since the angles here are meaningful only up to direction reversals, we take the smaller angle (in the [0,90] degree range) between the two corresponding lines (one corresponds to the curve normal and the other to the gradient direction). The ratio between the conditional likelihoods of the fold/cut cue on the right-hand side of a curve is plotted in Fig. 3(c). The tendency of a curve side with intensity gradients parallel to the curve to be 'figure' is apparent.

Parallelism cue. The parallelism cue [11] is based on the tendency of thin image segments to correspond to thin objects and not to narrow gaps between objects. This cue per curve C_i is computed as follows: For each pixel p on the curve, the closest curve pixel $p\prime$ (possibly on a different curve) is sought along the curve normal at p in the left direction. A similar search is performed in the right direction. Denote the distance between p and $p\prime$ by $r_{i_{\text{left}}}(p)$ for the left side and by $r_{i_{\text{right}}}(p)$ for the right side. Denote the difference between the angle of the curve tangent at p and the angle of the curve tangent at $p\prime$ by $\delta_{i_{\text{left}}}(p)$ for the left side and by $\delta_{i_{\text{right}}}(p)$ for the right side (as before, we take the smaller angle between the two tangents, which is in the [0,90] degrees range). Then, for each of the curve sides a "parallelism proxy" in [0,1] is computed:

$$\overline{s}_{i_{\text{left}}} = \frac{1}{|C_i|}\sum_{p=1}^{|C_i|} s_{i_{\text{left}}}(p), \quad (7)$$

where

$$s_{i_{\text{left}}}(p) = \max\left\{1 - \frac{r_{i_{\text{left}}}(p)}{|C_i|/2} - \frac{\delta_{i_{\text{left}}}(p)}{\pi/6}, 0\right\}, \quad (8)$$

and similarly for $\overline{s}_{i_{\text{right}}}$. The parallelism proxy per curve side is the mean score per pixel for that side, where the score per pixel per curve side is 1 if $\delta = r = 0$ and it decreases linearly in δ and in r until it reaches 0. These linear rates of decrease were chosen such that the maximal r and δ allowed are half the curve length and 30 degrees, respectively. Finally, the parallelism cue is the difference between the parallelism proxies for the two curve sides: $\overline{s}_{i_{\text{right}}} - \overline{s}_{i_{\text{left}}}$. The log-ratio of the conditional likelihoods of the cue is plotted in Fig. 3(d). As expected, the curve side containing close and approximately parallel curves tends to be perceived as figure.

3.2. Junction-related potential functions

As noted, each potential function V^y in (1) models the depth relations between the three image segments (S_{y_1}, S_{y_2}, S_{y_3}) sharing T-junction y by using cues related to this junction. Similarly to the pairwise potential functions V_i, the potential functions V^y are defined as the negative log posterior of the corresponding three segments' ordinal depths conditioned on the junction's angles, weighted by the length l_y of the shortest curve out of the three:

$$V^y\left(d_{y_1}, d_{y_2}, d_{y_3}\right) = -l_y \log \Pr\left(d_{y_1}, d_{y_2}, d_{y_3} | \theta_1^y, \theta_2^y, \theta_3^y\right), \tag{9}$$

where θ_i^y denotes the angle corresponding to segment y_i at junction y. Since the junction angles provide only indication regarding the relative depth ordering of the corresponding segments,

$$\begin{aligned} &\Pr\left(d_{y_1}, d_{y_2}, d_{y_3} | \theta_1^y, \theta_2^y, \theta_3^y\right) \\ &\quad \propto \Pr\left(\text{depth order of segments } y_1, y_2, y_3 | \theta_1^y, \theta_2^y, \theta_3^y\right). \end{aligned} \tag{10}$$

(Each event – three segment depths – in the PDF on the left-hand side is associated with one of the six events – depth orderings – on the PDF in the right-hand side. The three depths are mutually distinct by assumption.) Since the prior probabilities $\Pr\left(\text{depth order of segments } y_1, y_2, y_3\right)$ are equal for all six possibilities, the posterior is proportional to the likelihood of the angles:

$$\begin{aligned} &\Pr\left(\text{depth order of segments } y_1, y_2, y_3 | \theta_1^y, \theta_2^y, \theta_3^y\right) \\ &\propto p\left(\theta_1^y, \theta_2^y, \theta_3^y | \text{depth order of segments } y_1, y_2, y_3\right). \end{aligned} \tag{11}$$

Like the curve cues' conditional distributions, the conditional distribution of the T-junction angles (conditioned on the depth ordering of the junction segments) are non-parametric distributions learnt from the BSDS. A simple, but crucial observation is that although the annotation consists of only the binary information of which of the two curve sides is in front of the other, the depth ordering of the segments in a T-junction can be trivially inferred from this information. We also note that since the mutual PDF of $\theta_1^y, \theta_2^y, \theta_3^y$ is degenerate ($\theta_1^y + \theta_2^y + \theta_3^y \equiv 2\pi$), we learn the 2D conditional PDF of the two angles corresponding to the segment in the front and the segment in the back at a T-junction. This 2D distribution is shown in Fig. 4. The tendency of a segment with an approximately straight angle be perceived as the front one is apparent.

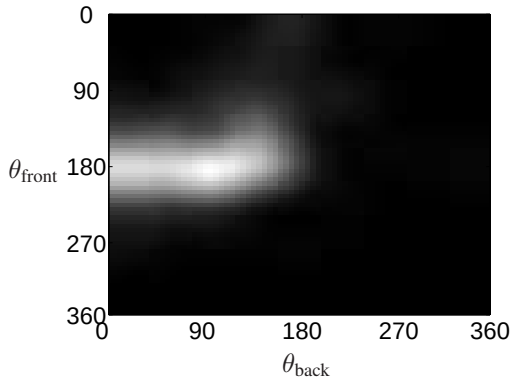


Figure 4. PDF of the two angles (in degrees) corresponding to the segment in the front (θ_{front}) and the segment in the back (θ_{back}) at a T-junction.

3.3. Potential functions related to non-adjacent segment pairs

As mentioned, each potential function $V_{j,k}$ in (1) models the depth relation between the two non-adjacent segments S_j and S_k using cues related to this segment pair. In our implementation, we modeled these functions using the observation that color distribution similarity between two image regions raises their likelihood of belonging to the same object, and thus also their likelihood of being of the same ordinal depth. Therefore, we include in the neighborhood system of the CRF (1) the non-adjacent image segment (unordered) pairs $\mathcal{P}$ where the hue distribution near one of the enclosing curves of one segment is very similar to the hue distribution near one of the enclosing curves of the other segment. Hue distributions along two curve sides are considered similar in our implementation if the cyclic earth mover's distance between the corresponding 64-bin normalized histograms (excluding pixels of saturation or value smaller than 15% of the total range [3]) is below 0.005. The corresponding potential function was set to

$$V_{j,k} = \begin{cases} \alpha, & d_j \neq d_k, \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

Overall best results were obtained by setting α to infinity, that is, enforcing the assignment of the same ordinal depth to image segment pairs containing a curve side pair of similar hue distribution.

3.4. Optimization

Maximizing the PDF (1) is equivalent to minimizing the corresponding *energy function*, which is the negative of the exponent of the exponential function in (1). Unfortunately, finding the global minimum of energy functions is already NP-hard when the variables are binary and the functions consist of only pairwise potential functions [7]. Therefore, we chose to estimate the optimal depth ordering by seeking "spatially local" optima of the energy function. In this optimization, the labels of each set of curves enclosing an image segment were estimated separately. Let the segments and the curves in the image be represented by a graph: each image segment is represented by a unique node where node adjacency represents segment adjacency. The labels of the curves enclosing an image segment were estimated by considering only the image segments and curves that are represented by the subgraph consisting of all nodes of distance d or less from the node that represents the former, enclosed image segment (hereinafter, "center segment"). This "spatially local" energy function was evaluated for all possible depth orderings, and the local depth ordering of smallest energy was chosen. This spatially local optimum of the depth ordering induced labels of the curves enclosing the center segment.

d was set such that the number of nodes in the subgraphs will be small enough for the full enumerations to take a reasonable time. For images consisting of M segments or less ($M = 9$ in our implementation), we set $d = \infty$, which reduced the above procedure to the full enumeration of all (global) depth orderings. Luckily, it turned out that this full enumeration was applied on about two thirds of the human-marked benchmark images. For the rest of the images d was set to 1, and the local enumerations were performed for the subgraphs consisting of M nodes or less. For a curve whose label could not be estimated by this method (due to too many image segments adjacent to the image segments on both its sides), the corresponding label was estimated according to the curves cues alone, that is, by maximizing (4). However, such cases were very rare (less than half a percent of all curve pixels). Note that two label estimates for a curve may be produced – one when the curve's left image segment was the center segment and one when its right image segment was the center segment. In cases where these estimates were conflicting, the curve's label was marked as undecided and was considered as half-correct/half-incorrect when labeling accuracy was evaluated.

The optimization runtime in MATLAB on a standard laptop computer ranged between a fraction of a second to about half a minute per image (depending on the number of image segments), with a mean of about 15 seconds.

We also tried simulated annealing and variants of it instead of the spatially local enumeration method above. These methods provided similar results, but were slower.

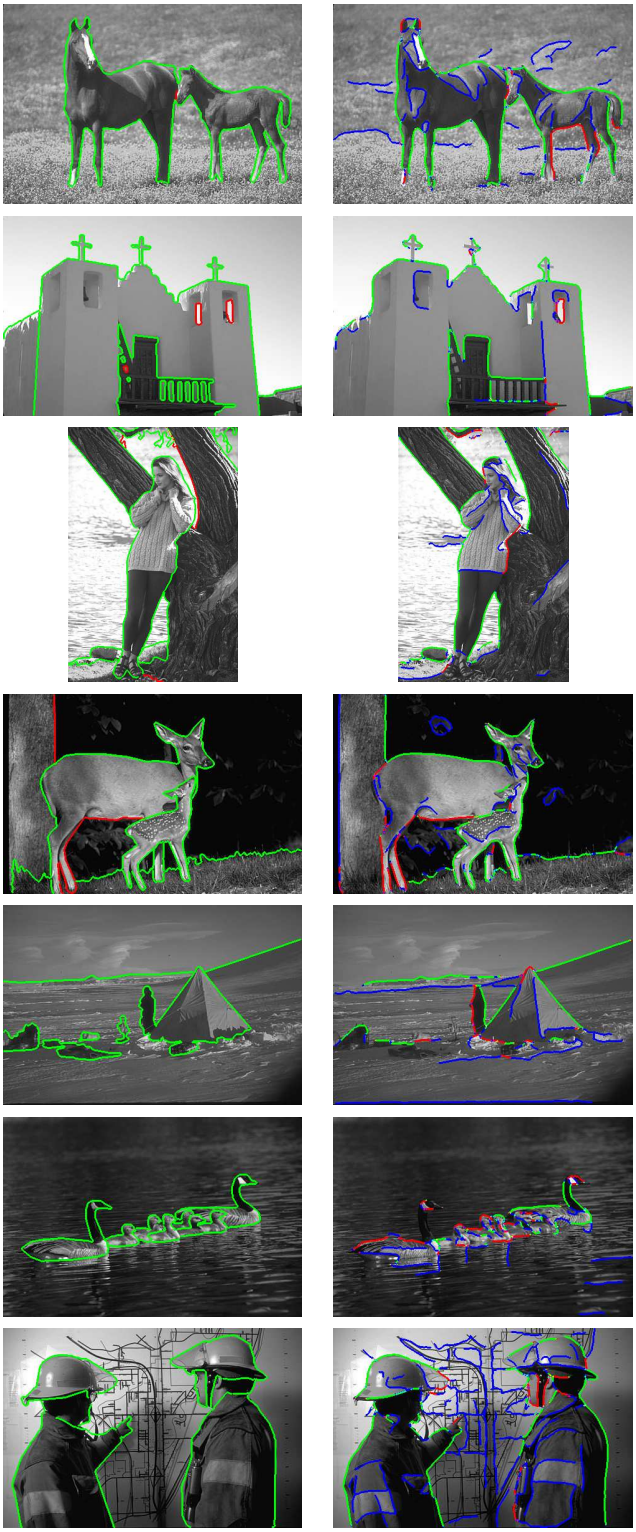
4. Experiments

The proposed method was tested using the same 100 test images used in [15], which were taken from the BSDS (the image list was taken from [12]). As mentioned, these images were first segmented by a human observer, after which two additional human observers attributed each of the boundary curves between the segments to one of its two abutting segments. As in [15], where the curve junction structure came from the human-marked segmentation, here the partitioning of the image into image segments uses the human-marked segmentation as well. As in [15], surface markings and curves labeled inconsistently by the two human observers were excluded from the experiments. All the PDFs of the curve cues (5), as well as the PDF of the angles in a T-junction (11), were learnt from the aforementioned 100 images using the leave-one-out cross-validation method. The corresponding figures in Sec. 3.1-3.2 show these PDFs (all 100 PDF sets look similar). Sample results are shown in Fig. 5(a).

As in [15], we evaluate the performance by counting the percentage of correctly labeled curve pixels. The best previously reported performance for the BSDS was the best performance reported in [15]: 78.3%, obtained for the CRF with respect to the curve junction graph ("Global CRF"). The non-parenthesized figures in Table 1 report the results of the proposed method for different cue combinations. (Note that the (non-parenthesized) figures in the first four rows report the performance when using only the corresponding curve cue, that is, only the V_i potential functions are included in the CRF (1) with only the corresponding cue PDF in (5).) As can be seen, when using all cues, the proposed method's accuracy is 82.8%, which is 20.7% fewer errors than reported in [15].

The parenthesized figures in the table report the performance when the curves are independently assigned the label of higher likelihood conditioned on the corresponding curve cue(s), that is, when the 2.1D model is disregarded. The accuracy gains by moving from independently assigning the curve labels to using the 2.1D model are also prominent. Another interesting observation from Table 1 is that the convexity and fold/cut cues seem to be more informative than the lower region and parallelism cues.

The previous experiment tested the proposed method on a benchmark of human-marked segmentations, where it yielded more accurate results than those previously reported. We did not have a way to reliably compare the performance of the proposed method to that of previous methods on automatically generated boundary curves. Nevertheless, we tested the method on automatically generated curves, using the same 100 BSDS images with their edge curves automatically generated using methods similar to those used to generate the curves in [15] (via [12]). In order to obtain the image segments, we connected each



(a) (b)

Figure 5. Sample results for (a) human-marked curves, and (b) automatically generated curves. Green denotes correct, red denotes incorrect, and blue denotes unmatched curve pixels. The original color images are shown here in gray-scale so that the curves can be seen clearly.

Cues	Performance
Convexity	71.4% (68.4%)
Lower region	64.1% (61.9%)
Fold/cut	71.8% (69.2%)
Parallelism	64.7% (52.4%)
Curve (V_i)	80.7% (78.1%)
Junction (V^y)	70.2%
Curve + junction	82.1%
All	**82.8%**

Table 1. Performance evaluation on the human-marked segmentations for various cue combinations. The non-parenthesized figures report the performance under the proposed model (2.1D). The parenthesized figures report the performance when the curves are *independently* assigned the label of higher likelihood conditioned on the corresponding curve cue(s).

curve end to the curve end closest to it or to its closest point on the image boundary, whichever was closer. A similar heuristic was used in [15] for constructing the junction graph. In order to transfer the ground truth labels from the human-marked curves to the automatically generated ones, we matched each pixel on the latter curves to its closest pixel on the former curves, while allowing a maximal Euclidean distance between matched pixels (0.75% of the image diagonal, which is about 4.3 pixels). Then, we transferred the ground truth labels according to the local curve orientations at the corresponding pixels[1]. Sample results are shown in Fig. 5(b).

As in [15], we measured the figure/ground assignment accuracy as the ratio between the number of correctly labeled pixels to the total number of pixels for which the ground truth label was transferred. This yielded 69.1% accuracy. Although this accuracy percentage for the automatically generated curves is similar to that in [15] (68.9%), the two methods might not perform similarly for such curves because of differences between the generated curves here and in [15] and because of the noted difference in the pixel matching process.

5. Discussion

A new method for estimating the figure/ground labels of boundary curves in images was proposed. The main novelty of the method lies in its use of the 2.1D model. The method estimates the ordinal depths of the image segments

[1] A similar process was carried out in [15], except that the matching between the pixels was bipartite. Bipartite matching between ground truth data and noisy data is more suitable under the condition that each ground truth datum generates at most one noisy datum. Looking at the ground truth and generated edge maps, we concluded that this condition does not hold and so we opted for the other matching method. Note also that a bipartite matching might match only a fraction of a noisy tortuous curve to its corresponding true, smooth curve. This will reduce the weight of the curve in the overall accuracy evaluation, which might bias the evaluation positively since the labeling of noisy curves is more likely to be wrong.

enclosed by the curves and then uses this estimate to infer the figure/ground labels of the curves, rather than estimating the figure/ground labels directly. This approach eliminates from consideration figure/ground label configurations that do not match any 2.1D model and are thus very likely to be incorrect. Another advantage of the approach is that it enables the incorporation of cues related to image segments. These cues may also relate non-adjacent segments. There does not seem to be a natural way to incorporate such cues when modeling the figure/ground curve labels directly. All the curve cues and the junction cues are incorporated into the probabilistic model by using their non-parametric conditional distributions, which are learnt from a human-marked training set. The implementation of the approach produced figure/ground assignments for human-marked BSDS images with 20.7% fewer errors than the state-of-the-art for the same benchmark.

Note that the assumption that image segments separated by a curve lie on different depths is made here for comparison to [15], where this assumption was made as well. Adjusting the proposed approach so that adjacent segments can be assigned the same depth requires only simple changes: that the different cue PDFs be learnt conditioned on this case (in addition to the "different depths" cases, which are already learnt), and that the prior probabilities of this case for curves and junctions be learnt as well. The optimization, of course, would then be over all depth orderings and not restricted to depth permutations.

All methods for assigning figure/ground labels to curves, including the proposed one, make use only of cues related to the curves themselves and to the image in the close neighborhood of the curves. It would be interesting to know the limits of approaches that are restricted to using only such information. For example, consider the images in Fig. 6. These are four images from the test set with all but the neighborhood around the human-marked curves concealed. It appears that a human observer would perform poorly on the boundary ownership task when exposed to this version of these images alone. If we assume human performance to be an upper limit on the performance of an algorithm for this task, this gives us some indication as to how the best algorithm can be expected to perform under the same circumstances. Experiments on human observers were conducted in [1] to assess how much information about figure/ground assignment is available from locally computed cues. Images such as those in Fig. 6 can be used for generalizing these experiments to assess the information available from these cues globally.

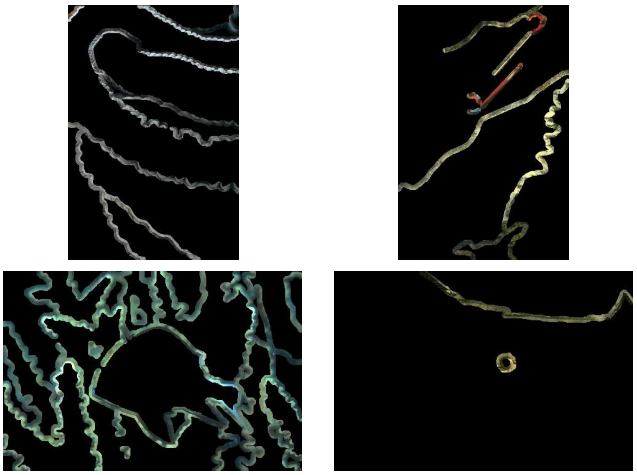


Figure 6. Four BSDS images from the test set with all but the neighborhood around the human-marked curves concealed. For what percentage of the curves can you decide on the boundary ownership correctly? See text for details.

References

[1] C. Fowlkes, D. Martin, and J. Malik. Local figure-ground cues are valid for natural images. *Journal of Vision*, 7(8):2:1–9, 2007.

[2] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *PAMI*, 6(6):721–741, 1984.

[3] T. Gevers and A. Smeulders. Color based object recognition. *Pattern Recognition*, 32:453–464, 1999.

[4] D. Hoiem, A. Stein, A. Efros, and M. Hebert. Recovering occlusion boundaries from a single image. *ICCV*, 2007.

[5] P. Huggins, H. Chen, P. Belhumeur, and S. Zucker. Finding folds: On the appearance and identification of occlusion. *CVPR*, 2:718–725, 2001.

[6] G. Kanizsa and W. Gerbino. Convexity and symmetry in figure-ground organization. In *Vision and Artifact*, pages 25–32. New York: Springer, 1976.

[7] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *PAMI*, 26(2):147–159, 2004.

[8] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *ICML*, pages 282–289, 2001.

[9] G. Mori, S. Belongie, and J. Malik. Shape contexts enable efficient retrieval of similar shapes. *CVPR*, 1:723–730, 2001.

[10] M. Nitzberg and D. Mumford. The 2.1-D sketch. *ICCV*, pages 138–144, 1990.

[11] S. Palmer. *Vision Science - From Photons to Phenomenology*. MIT Press, 1999.

[12] X. Ren. Private communication.

[13] X. Ren, C. Fowlkes, and J. Malik. Cue integration for figure/ground labeling. *NIPS*, 2005.

[14] X. Ren, C. Fowlkes, and J. Malik. Scale-invariant contour completion using conditional random fields. *ICCV*, 2:1214–1221, 2005.

[15] X. Ren, C. Fowlkes, and J. Malik. Figure/ground assignment in natural images. *ECCV*, 2:614–627, 2006.

[16] S. Vecera, E. Vogel, and G. Woodman. Lower region: A new cue for figure-ground assignment. *Journal of Experimental Psychology: General*, 131:194–205, 2002.